Computational observer approach for the assessment of stereoscopic visualizations for 3D medical images

Fahad Zafar, John Dorband and Aldo Badano

Center for Devices and Radiological Health, FDA, Silver Spring, MD.

1. ABSTRACT

As stereoscopic display devices becomes common, most performance studies conducted on these 3D displays are based on either physical measurement of display characteristics using instrumentation or psycho-physics experiments with humans quantifying their experience. We present a computational stereoscopic observer approach inspired by the mechanisms of stereopsis in human vision that makes decisions based on a set of image pairs. Our stereo-observer is constrained to a left and a right image generated using a visualization operator (ray tracing) to render simulated voxel data sets. We present the formulation of the observer based on model observer theory and discuss issues regarding simulated data generation and processing for this approach. The applicability of this observer extends to stereoscopic displays in the areas of entertainment, industrial, and medical imaging applications.

2. INTRODUCTION

Display technologies have seen tremendous innovations in the past few years. Stereoscopic 3D display devices are now being mass marketed to consumers for entertainment and research applications alike.[1] As a result quantitative analysis of the performance of these devices has become increasingly important.[2,3] Most studies conducted with 3D displays work with human subject[4–7] and include feedback from the human observer where they report the experience after viewing the content or differences in particularly tailored 3D content. Though few results have been somewhat consistent, there is no quantitative data that shows how much of these factors are tolerable and to what extent. Out of a large number of existing human observer studies only a very few of them are related to medical imaging. Most of the content used comes from entertainment and content generated for video games. Thus, a rigorous study on the impact of 3D display technology in medical imaging diagnosis is required. With the current 3D display technology trends and new products in the market, it is not feasible to conduct a human observer study for each and every device before it is approved for field use. Recently model observer approaches have been proposed and used to demonstrate performance levels of medical imaging devices submitted to the FDA. Statistical decision theory presents the concept of a decision maker, or model observers that are used to analyze large data sets which help in assessment of image quality. We intend to model the human observer by using a stereo observer that will mimic its characteristics and explore the issues in 3D stereo display systems including crosstalk, noise and hardware limitations (active vs passive). Image quality deterioration by one 3D technology can be simulated by applying a distortion function in the spacial/frequency domain, which will subsequently affect the final results generated by the observer. This way we can study and understand many issues which would otherwise require a human observer study with likely mixed, limited and sometimes unreliable data.

The objective evaluation of imaging systems requires three components as stated by *Barret and Myers* [8].

1. Task: Definition of the use of images.
2. Observer: Makes use of the images in order to make a decision.
3. Figure of Merit: Standard metric used for evaluating performance.

Medical imaging tasks are generally categorized as classification or estimation. Classification is a binary decision process where the presence or absence of the signal is determined. Estimation, is much more interesting since it involves quantitative description of the object based on the data. In this work we present our initial layout of the experiments for signal detection tasks. Later, we formally define the stereo observer, the data set generation and explain our findings with data set generation methods and the modeling of the problem.

2.1 Signal Detection

In signal detection tasks, the observer needs to classify images as either with signal or without signal. The decision is based on a threshold generated using statistical

From the Division of Imaging and Applied Mathematics, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, U.S. Food and Drug Administration. 10903 New Hampshire Ave., Building 62, Room 3116, Silver Spring, MD 20993, phone 301-796-2534, email aldo.badano@fda.hhs.gov.

properties of the input data. 9 presented the vCHO, an observer that views the data across the sagittal plane where each noise value generated using a Gaussian noise functions directly represent pixels within the images. Additionally, the image statistics generated retain the properties of the original Gaussian Noise function (*such as mean and variance*). The advantage of stereo vision is the additional queues for depth perception provided by two views instead of a single one. The stereo observer thus has more data to work with and should be able to make a more informed decision. In order to generate stereo pairs we need to represent each noise value as a voxel rather than a pixel. Unless voxels are used to generate stereo images, no new information will be contained with in the pair. Simple data shifting techniques used for generating Random Dot Stereograms (RDS) will not provide a visualization system that is capable of generating a robust stereo data set. The problem of retaining the original Gaussian noise characteristics with in the data tends to get difficult. The reason being that voxels need to be transparent in order to generate projections with depth information. Each slice in the stack now contributes to the image since the slices in front of it have some transparency. This additional issue is not present for a 2D or 3D observer. Additionally, some additional artefacts from the imaging system tend to emerge since one voxel might cover more than a single pixel. This would alter the covariance matrix for the image since gaussian noise values no longer contribute directly to one pixel.

3. STEREO IMAGING

A general digital imaging system as described by[10] involves the collection of a discrete set of data, represented by vector g from a continuous object f(r,t). The system operator which maps the continuous object to discrete data space is represented by H.

$$g = Hf + n \tag{1}$$

It should be noted here that once g is obtained, it is incorporated directly into the observer models presented by.[9] This assumption is not practical for an anthropomorphic observer since human observers consume g though a visualization technique applied to correctly visualize the voxel data using a display device. This operator S_g (Eq. 2) is the transfer function that brings the medical dataset from reconstructed space to screen space. In our representation, it is dependent on the transparency set for each slice of voxels with in the stack. Other parameters such as eye seperation and voxel size have been kept constant for this research. There is yet another visual limitation that is directly related to the luminance characteristics of the display device S_v. The luminance to graylevel mapping of the device will eventually affect the final rendered images presented to the human observer. This function can help analyze limitations of crosstalk that can be tolerated with in a display system. Crosstalk is defined as signal transmitted on one circuit causing undesirable effects in another. In 3D display systems, crosstalk means the contamination of one eye image into the other. In terms of vision, it can be stated that the luminance that was intended for one eye is bleeding into the luminance of the other. S_v is one such operator that is related directly to the display system and this is where we incorporate our crosstalk function to deteriorate the image signal for each eye. In active 3D display systems, the shutter glasses refresh each eye image at 120Hz each. Each eye experiences full access to the light signal followed by complete blackout in traditional interleaved manner. So, 100 percent crosstalk is experienced by the user when he stares at an active display screen without the shutter glasses. Due to the manufacturing and technology limitations, there is always some measure of crosstalk for every system. The idea is to keep it as low as possible so as to keep this ghosting effect experienced by the observer to a minimum. For instance, we will introduce some percentage (β) of crosstalk into the system and calculate the SNR to understand its effect Eq.3, Eq. 4. Some device specific parameters (*luminance, resolution, refresh rate*) would be needed in order to particularly tailor S_v for a specific display. For this discussion we choose to ignore device specific characteristics and their effects on the model but it remains as part of our future direction.

$$g \xrightarrow{S_g\{\alpha\}} (g_l, g_r) \xrightarrow{S_v} (L_l, L_r) \tag{2}$$

where S_g is the visualization operator, S_v is the display device post processing functions, g_l, g_r stereo-pair of left and right eye. L_l, L_r luminance through display for left and right eye, α alpha/transparency.

$$ImageLeft_{i,j} = ImageRight * \beta + ImageLeft(i,j) \tag{3}$$

$$ImageRight_{i,j} = ImageLeft * \beta + ImageRight(i,j) \tag{4}$$

where: β is the percentage of crosstalk, (i,j) : position of each pixel in the images

We define g_l and g_r as the actual data viewed after rendering using ray-box intersection technique presented by.[11] The complete description of the method is presented by[12] and.[13] After the stereo views generation for the scene, we apply a post processing function S_v depending on the display device parameters. This final results obtained is the luminance left (L_l) and luminance right (L_r) that is fed into the stereo observer. The sensitivity of the experiment is dependent on the functions S_g and S_v.

3.1 Ideal Linear Stereo Observer

Statistical decision theory presents the concept of a decision maker, or observer that determines which of the two classes the data set belongs to. The Ideal Observer (IO) is one that makes use of all available information in both the images as well a prior information, to optimize classification performance. Full knowledge pertains to the conditional probability density function of the image data under each hypothesis H_1 and H_2

$$H_1 : (g) = b + s \tag{5}$$

$$H_2 : (g) = b \tag{6}$$

where b: background (includes noise), s: signal).

The test statistic of the IO under the light of the existing hypothesis (H_1 and H_2) is defined as follows:

$$\Delta(g) = \frac{pr(g|H_2)}{pr(g|H_1)} \tag{7}$$

Most of the times, we may not be able to evaluate the IO performance simple because the probability density functions are unknown. Secondly, we might be interested in a prediction task as assessed by humans. In this case, we move to assess the performance using the Stereo Ideal Linear Observer. When the ensemble mean and data covariance for g is known, the observer that maximizes the test statistic (*SNR in our case*) can be derived and is known as Hotelling Observer or the Ideal Linear Observer. We define L_g as the serialized concatenation of left and right luminance emitted from the display device and present the formulation of our Ideal Linear Stereo Observer (ILSO).

$$L_{\overline{g}} = L_l \oplus L_r \tag{8}$$

$$SNR^2 = s^t {K_{L_{\overline{g}}}}^{-1} s \tag{9}$$

where s: signal, g: given image), s^t: signal transpose, K_{L_g}: covariance matrix for particular background.

g, is the image whose entries g_m, where m= 1;...;M, are the intensity of image pixels in 2D data and M is the number of elements present in the image. An observer is defined by its discriminant function, which maps an image g to its test statistic t = *t(g)*. The decision is made by comparing the test statistic to a certain threshold t_0. When $t > t_0$, the signal is present, hence H_1 holds, and the image is classified as signal-present. Otherwise, H_2 is satisfied and the image is classified as one with signal-absent.

In the case of stereo-imaging, the observer has access to a pair of images, two projections through the 3D object visualized using volume rendering (S_g). Once we obtain a single stereo pair, g is defined as a vector containing all the pixels in the left and the right view (Eq. 8). S_g is the ray tracing operator that outputs stereo images rendered through off-axis projection. Steps involved when applying S_g to the system are explained in detail in the later sections. S_v is a set of post processing operations applied to the images based on display characteristics. In our current experimental setup S_v is an Identity Matrix (I) that essentially has no effect on the output.

Our approach is to aim for an anthropomorphic observer when dealing with stereo vision. Through a variety of postulated mechanisms,[14] the human visual system reconstructs the 3D object. An observer interested in signal estimation would therefore perform the reconstruction of the object using the information of the stereo-image pair or at least an estimate for the signal depth. For now however, we intend to quantify that stereo image pairs present more information compared to a single 2D image viewing system in signal detection tasks. The 3D channelized Hotelling observer[9] works with 2D images without incorporating the imaging system (S_g, S_v). In light of our problem definition our results will not be comparable unless Gaussian noise characteristics are are kept constant through S_g, which we find is not the case. We need to generate an SNR for a 2D Ideal Linear Observer through the imaging system (S_v) and compare it to our stereo observer performance. This comparison will show in essence that stereo image viewing increases the information gain when visualizing medical data. Once we establish and verify this baseline stereo observer, we can move to advance the sensitivity of our stereo observer.

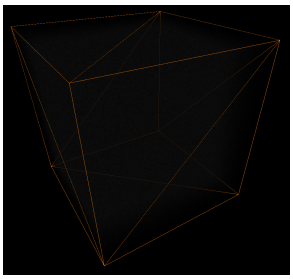
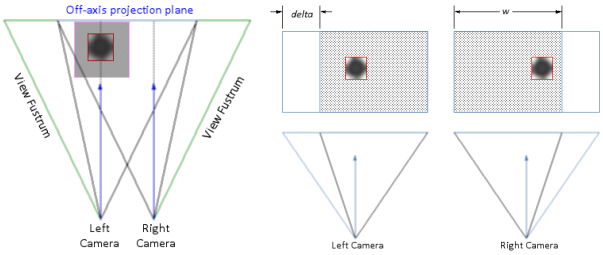


Figure 1. Volume Rendered noise cube with bounding box added for perspective



(a) (b)

Figure 2. (a) Stereo pair generated when both cameras projecting on the off-axis projection plane. Signal displayed in red border, noise cube in purple border. (b) Trim delta amount of pixels from left side of (g_l) and right side of g_r. Signal displayed in red border.

4. METHODS

Once the stereo observer is formulated, it requires a data set in order to calculate performance. The data set necessary for calculating the test statistic either comes from an imaging system or is simulated depending on the desired background. In this work we present our findings with white noise and leave lumpy and cluster lumpy backgrounds for future research. We start by developing a simple chain of steps that will allow for experimentation and testing. Our stereo data generation requires additional stages of processing when compared to traditional simulated image generation for model observers. In this section we explain the functional aspects of our data generation techniques.

4.1 Ray Tracing

Volume rendering is a well known technique used to render 3D discretely sampled voxelized data. We generate a noise cube sized at 256x256x256. Then each voxel is assigned a random noise value. Our implementation of White noise uses the the Mersenne Twister Random Number Generator. We do not apply techniques (such as marching cube) that extract isosurfaces from the data. Rather we use ray tracing using Ray-Box intersection[13] technique. This methodology is well studied and understood and documented.[12,13] Once the data set is generated the opacity is used to define the color of each voxel. We do not use colored data sets, thus a simple *transfer function* is used that maps the value of the voxel directly to its opacity in graylevel. A transfer function is used to map voxel values to Red, Blue, Green and Alpha (RGBA) values. Color in graphics is represented as a four component vector where the last component alpha stands for the opacity. Alpha blending is a technique used to mimic semi-transparent surfaces by blending a series of surfaces till the resultant alpha of the final color component is equal to 1. For real world objects transfer functions are well defined in order to bring out certain features of the objects (eg. bone, muscle). For our simulated noise data set, we employ a transfer function that multiplies the gray level value to a fixed alpha (0.05 in our case on a 0-1 scale). For about a 1000 iterations of the ray through the volume, at each step we integrate the new voxel the ray hits with the previous color value the ray has accumulated along its directed path. Note that the alpha value directly affects the overall brightness of the data set, the lower the alpha, the more transparent the volume. This value should be set based on the number of steps the ray takes forward through the volume. Figure 1 shows the rendered noise cube as output by the system.

4.2 Simulated Data Set Generation

When White noise with N(0,1) is generated, these values cannot directly be display using an RGB color system. Though GPUs treat color as a 4 unit vector ranged from 0-1 for each RGB and A (alpha), they cannot handle negative values to display gray levels. We shift the range of our noise cube to N(128,32) *(mean=128, stdev=32)* in order to display the simulated data. White noise as a background is well understood and commonly used with in medical imaging literature.[9] Though we theoretically know how the covariance matric for such a background would look like (*a diagonal matrix with the noise variance*), we need to recalculate the covariance matrix for this background after applying the imaging system. This covariance is used to calculate the Signal to Noise Ratio (SNR) for the Ideal Linear Observer.

4.3 Stereo Pair Generation

Stereo image generation comes natively when using rendering libraries such as OpenGL[15] and DirectX.[16] Simple commands inserted with in the code structure will render two views with some preset parameters. Though this method is extensively employed for video games and entertainment content, it grants limited access to

changing the variables associated with rendering stereo pairs. Some driver level access to such parameters can be achieved using the NVAPI[17] which allows direct access to NVIDIA GPUs and drivers but these programming paradigms tend to be very low level architecture based environments. In our stereo visualization system, we render volumes using XNA4.0[18] which is built on top of the DirectX API. The stereo image generation is a separate component that uses off-axis projection to generate the pair of images. This helps us visualize our volumes in time-multiplexed 3D displays[19] (*active-3D*) as well as Planar's Stereomirror.[20, 21]

The concept of stereo images generation in our system is presented in Figure 2. Figure 2 (a) presents the top view of the system where we see the nosie cube with an embedded signal and the camera positions. Both cameras need to render the image on the same plane using an off-centered projection. Once these images are available, the last step is to trim the correct amount of pixels (δ) from both the images to obtain a stereo pair (Figure 2). Note that there are many faults that can creep in when generating stereo pairs regarding positive parallax or negative parallax or vertical parallax (if the images are not rendered using off-axis projection). These concerns have been dealt with and our rendered images are free from these rendering issues.

$$\delta = \frac{e.w}{2.f.\tan(\theta/2)} \tag{10}$$

Where e: eye separation, w: intended width of the expected plane, f: focal distance, θ: camera aperture or field of view

Since we use software cameras to render the g_l and g_r (*the stereo image pair*), all these parameters are known and can be changed easily for a range of experiments. The projection matrix used for the software cameras is as follows:

$$\begin{bmatrix} xScale & 0 & 0 & 0 \\ 0 & yScale & 0 & 0 \\ 0 & 0 & farZ/(farZ - nearZ) & 1 \\ 0 & 0 & -nearZ.farZ/(farZ - nearZ) & 0 \end{bmatrix}$$

$$yScale = \cot(\alpha/2) \tag{11}$$

$$xScale = yScale/AspectRatio \tag{12}$$

where $farZ$: distance to the far view plane, $nearZ$: distance to the near view plane, α: field of view in the y direction, in radians, $AspectRatio$: aspect ratio of the viewport.

A ray casting GPU program (*shader*) is used to trace rays across a noise cube volume. We plan to move the implementation of our entire experimental chain to CUDA in order to decrease data set generation and covariance calculation times.

5. RESULTS AND DISCUSSION

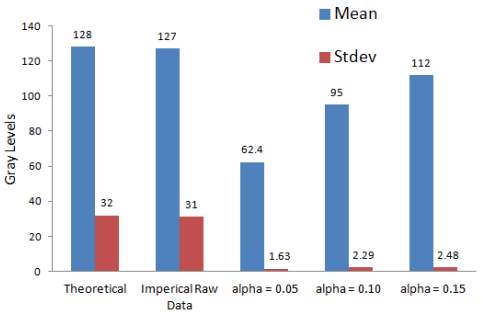


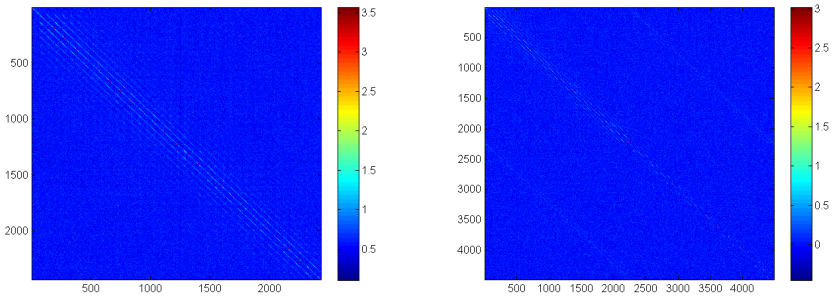
Figure 3. Mean and variance for the data set before and after applying visualization

We intended to calculate the SNR for a stereo observer and compare it to the 2D observer from existing literature. It is our finding that it is an unfair comparison since the data variance drops substantially once S_g is introduced (Fig.3) in the system. Since the variance has decreased, the SNR tends to be very high for the stereo observer as compared to a 2D observer. Variance for white noise is preserved with in the raw data (Fig.3) but once it is processed for screen space, its attributes tend to change. This is the trait expect when any imaging operator is applied to an object. In our case S_g is necessary to generate the simulated stereo projections. Additional artefacts also appear due to the imaging system since each noise point is represented as a voxel (in 3D point space) rather than a 2D pixel. Voxel size, camera angle and its positioning are other factors to consider when simulating a projection system to generate stereo views. These tends to have an affect on the properties of g. It is our understanding that we compare the left/right only image data with the stereo data. We would calculate the covariance separately for left/right images and the stereo pairs. These covariance matrices will factor into the SNR equation to present observer performance results on the same scale. The alternative direction would be to continue exploring the effect of imaging system factors to bring the final variance of the visualized data to as close as possible to the theoretical variance. We have already shown how

We believe our initial findings are encouraging. We notice the additional correlation in the covariance for the stereo data where the left image pixels have a weak correlation with the right image pixels. This additional side diagonal is weaker than the major diagonal in the covariance matrix. This additional information has increased the detection likelihood for the signal but it has been clouded by the reduction in the variance of the data. This overshadowing of performance due to the imaging system makes it difficult to explore this additional left-right correlation information contained in the covariance matrix. We also intend to speed up our experimentation process by incorporating multi-core techniques. Data generation and covariance calculations are time consuming operations for large number of images. Currently, we use a single core implementation that takes about 8 hours for a complete run, including data generation. Our current simulated data set is sized at 7000 images.

REFERENCES

1. C. H. Messom. Abstract stereo vision controlled humanoid robot tool-kit, Massey University, Albany, New Zealand.
2. Jukka Hakkinen, Jari Takatalo, Markku Kilpelainen, Marja Salmimaa, and Gote Nyman. Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width. Journal of the Society for Information Display – May 2009 – Volume 17, Issue 5, pp. 433-441.
3. H. Liang, S.Park, B. D. Gallas, K. J. Myers, and A. Badano. Image browsing in slow medical liquid crystal displays. Acad. Radiol. 2008.
4. HP Chan, MM Goodsitt, MA Helvie, LM Hadjiiski, JT Lydick, MA Roubidoux, JE Bailey, A Nees, CE Blane, and B. Sahiner. Roc study of the effect of stereoscopic imaging on assessment of breast lesions. Med Phys., 2005.
5. Masakazu Funakoshi Kazunori Shidoji and Masahiko Ogawa. Perception of absolute and relative distances in stereoscopic image. Stereoscopic Displays and Applications XXI. SPIE, 2010.
6. Kuo-Chung Huang, Jinn-Cherng Yang, Chou-Lin Wu, Kuen Lee, and Sheue-Ling Hwang. System-crosstalk effect on stereopsis human factor study for 3d displays. Stereoscopic Displays and Applications XXI. SPIE, 2010.
7. J.P. Rolland and H.H. Barrett. Effect of random background inhomogeneity on observer detection performance. J. Opt. Soc. 1992.
8. Jacob Beutel, Harold L. Kundel, and Richard L. Van Metter. Handbook of medical imaging. volume 1. physics and psychophysics. SPIE Press 2000.
9. Ljiliana Platisa, Bart Goossens, Ewout Vanteenkiste, Subok Park, Brandon D. Gallas, Aldo Badano, and Wilfried Philips. Channelized hotelling observers for the assessment of volumetric data sets. Journal of Optical Society of America. Vol 28, No. 5 May 2011.
10. H. H. Barrett and K. Myers. Foundations of Image Science. John Wiley & Sons, 1st edition, 2004.
11. T. L. Kay and J.T. Kajuya. Ray tracing complex scenes. Computer Graphics, 20(4), pp. 269-78. 1986.
12. Andrew S. Glassner. An overview of ray tracing. Andrew Glassner, ed., Academic Press Limited, 1989.
13. SIGGRAPH. Ray - box intersection. www.siggraph.org/education/materials/hypergraph/raytracer/rtinter3.htm.
14. Heather R. Filippini and Martin S. Banks. Limits of stereopsis explained by local cross-correlation. 9(1), 2009.
15. Mark Segal and Kurt Akeley. The opengl graphics system: A specification. http:www.opengl.orgdocumentationspecs version2.0glspec20.pdf, 2004.
16. Microsoft. Directx SDK. http:msdn.microsoft.comlibraryee663275%28VS.85%29.aspx, 2010.
17. NVIDIA. Nvidia driver settings programming guide. http:developer.download.nvidia.comassets toolsdocsPG-5116-001_v02_public.pdf, 2011.
18. Microsoft. What's new in xna game studio 4.0. http:msdn.microsoft.comen-uslibrarybb417503.aspx.
19. NVIDIA. Nvidia 3d vision automatic. http:developer.download.nvidia.comwhitepapers 2010NVIDIA%203D%20Vision%20Automatic.pdf, 2010.
20. D J Getty and P J Green. Clinical applications for stereoscopic 3-D displays. J. Soc. Inf. Displ., 23:36, 2007.
21. D. J. Getty, C. J. Dórsi, and R. M. Pickett. Stereoscopic digital mammography: Improved accuracy of lesion detection in breast cancer screening. Proceedings of the IWDM, pages 74–79, 2008.



(a) Left (g_l) (b) Stereo (L_g)

Figure 4. Visualizing covariance matrix for g_l and L_g